



SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Yanagisawa, Masashi
Bergsma, Derk
Wilson, Shelagh
Brooks, David
Gellai, Miklos

(ii) TITLE OF THE INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE
RECEPTOR HFGAN72

(iii) NUMBER OF SEQUENCES: 21

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: United States of America
(F) ZIP: 19406

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 08/938,548
(B) FILING DATE: 26-SEPT-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/887,382

(B) FILING DATE: 2-JUL-1997

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/820,519

(B) FILING DATE: 19-MAR-1997

(A) APPLICATION NUMBER: 60/033,604

(B) FILING DATE: 17-DEC-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: King, William T.

(B) REGISTRATION NUMBER: 30,954

(C) REFERENCE/DOCKET NUMBER: ATG50037-2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 610-270-5219

(B) TELEFAX: 610-270-4026

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1970 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAAACATAAT GTGGGTCTCG CGTCTGCCTC TCTCCCGCCC CTAATTAGCA GCTGCCTCCC	60
TCCATATTGT CCCAGGCCAG CGCTTCTTT GTGCTCCCAG ATTCCCTGGGT GCAAGGTGGC	120
CTCATTAGTG CCCGGAGACC GCCCCATCTC CAGGGAGCAG ATAGACAGAC AAGGGGGTGA	180

TCAGGGGCAC	AGTGATCCAA	CCCTGGCCTC	TGAACGCCGC	AGCGGCCATT	CCTTGGGCC	240
AGCCTGGAGA	CGGCCCGCC	GCAGCAGGCT	AATCTTAGAC	TTGCCTTGT	CTGGCCTGG	300
TGTGGACGCA	ATGTGCCTGT	CAATTCCCCG	CCACCTCAGA	GACTATAAA	CCCCAGACCC	360
CTGGGAGTGG	GTCACAATTG	ACAGCCTCAA	GGTTCCCTGGC	TTTTGAACC	ACCACAGACA	420
TCTCCTTCC	CGGCTACCCCC	ACCCTGAGCG	CCAGACACCA	TGAACCTTCC	TTCCACAAAG	480
GTAAAGATCC	AGGGATGGAG	GGGTGACTCA	GCCATCCCAG	AGGAAGCAAA	AAGAGTGTCTT	540
GCTCAGAGGG	CTGGAAGAAA	GGCCAAAGGT	GTCTCCACTC	TTGGTCTTTT	CCTGGGTGTG	600
CTCTGAGGCA	GGAGCACCTG	CCTTGGCTCA	CATTGGGTTG	GGTGCTGTTT	TGCTAAGAGC	660
CTGTGTTGC	TGAGCTCATA	TGTGTCAGGT	GCTCCGTTG	CACCTGTCAT	CTCTTGTCT	720
CCTCCCAACA	GCCTTGCAGA	GTAGAAATTA	TTTCTAGTAT	ACCCAGTTA	CAGGTAAGGG	780
AGCTGTGCC	TCTGAAAGGG	CAGGAAACTG	GTTCAAAGCA	ACGGAGTTCA	GTCACTCCTG	840
CAAGGGGGCA	GGCAGATGAG	AGAGCATTCT	GGAGTCTTGC	TAGTTCTGA	TTTCCATGTG	900
TTTCCCTGCT	GTGGAGAGGA	AGTTGGGGGG	ACTCAGTAGG	GCCCAGGTTT	TTCCCAAGTT	960
TACAACCTCT	GCTGCAGACA	GACACTCCTG	TTTCAGGTG	GAGTGGCAAG	TGCCCTAGTG	1020
GTGGCAACAG	TGGCCTAAGT	CTCCAGAGAA	AAGGGGGATT	CACTCTGCC	AGGGGGTCTC	1080
AAAAGGCTTC	CTGTGGGAGA	TGCTCTGCTG	GGTCTTGAAG	GAGGAGCAGG	GAAAGTAGGC	1140
CGATACCAGC	AAGGGCGCAA	AGCAAGGAGA	ACTAAGTGAC	AGCCAGAAAG	GAGTGCAGGC	1200
TTGGAGGGGG	CGCGGAGCCA	GAGGGGCAGG	TCCTGTGCGT	GGGAGCTGGT	GGCGGGGCC	1260
GTGGGAAGAC	CCCCCCAGCG	CCCTGTCTCC	GTCTCCCTAG	GTCTCCTGGG	CCGCCGTGAC	1320
GCTACTGCTG	CTGCTGCTGC	TGCTGCCGCC	CGCGCTGTTG	TCGTCCGGGG	CGGCTGCACA	1380
GCCCCCTGCC	GAUTGCTGTC	GTCAAAAGAC	TTGCTCTTGC	CGCCTCTACG	AGCTGCTGCA	1440
CGGCGCGGGC	AATCACCGCG	CCGGCATTCT	CACGCTGGGC	AAGCGGAGGT	CCGGGCCCCC	1500
GGGCCTCCAG	GGTCGGCTGC	AGCGCCTCCT	GCAGGCCAGC	GGCAACCACG	CCGCGGGCAT	1560
CCTGACCATG	GGCCGCCGCCG	CAGGCGCAGA	GCCAGCGCCG	CGCCCCCTGCC	TCGGGCGCCG	1620
CTGTTCCGCC	CCGGCCGCCG	CCTCCGTGCG	GCCCCGGAGGA	CAGTCCGGGA	TCTGAGTCGT	1680
TCTTCGGGCC	CTGTCCTGGC	CCAGGCCTCT	GCCCTCTGCC	CACCCAGCGT	CAGCCCCCAG	1740
AAAAAAAGGCA	ATAAAAGACGA	GTCTCCATTG	GTGTGACTGG	TCTCTGTTCC	TGTGCGGTG	1800
CGTCCTGCC	ATCCGGGGTG	GCAAAGCGTC	TTGCGGAGGA	CAGCTGGGCC	TGGAAGCCCG	1860
GCTGTCGGGC	ACCAGCCTTA	GCTTTGCGT	GGTTGAATCG	GAAACACTCT	TGGTTGGGGA	1920
GTTCCTCAGTG	CAAGGCCCTG	GGGCACAGAG	AGAACTGCAC	AGGTGCATGC		1970

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 131 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Leu Pro Ser Thr Lys Val Ser Trp Ala Ala Val Thr Leu Leu
1 5 10 15
Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Ser Gly Ala Ala
20 25 30
Ala Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg
35 40 45
Leu Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu
50 55 60
Thr Leu Gly Lys Arg Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu
65 70 75 80
Gln Arg Leu Leu Gln Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr
85 90 95
Met Gly Arg Arg Ala Gly Ala Glu Pro Ala Pro Arg Pro Cys Leu Gly
100 105 110
Arg Arg Cys Ser Ala Pro Ala Ala Ser Val Ala Pro Gly Gly Gln
115 120 125
Ser Gly Ile
130

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
1 5 10 15
Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
20 25 30
Leu

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln
1 5 10 15
Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr Met
20 25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGCTCGGCCGG	CCTCAGACTC	CTTGGGTATT	TGGACCACTG	CACCGAAGAT	ACCATCTCTC	60
CGGATTGCCT	CTCCCTGAGC	TCCAGACACC	ATGAACCTTC	CTTCTACAAA	GGTTCCCTGG	120
GCCGCCGTGA	CGCTGCTGCT	GCTGCTACTG	CTGCCGCCGG	CGCTGCTGTC	GCTTGGGGTG	180
GACGCGCAGC	CTCTGCCCGA	CTGCTGTCGC	CAGAAGACGT	GTTCCCTGCCG	TCTCTACGAA	240
CTGTTGCACG	GAGCTGGCAA	CCACGCCGCG	GGCATCCTCA	CTCTGGGAAA	GCGGCGACCT	300
GGACCCCCAG	GCCTCCAAGG	ACGGCTGCAG	CGCCTCCTTC	AGGCCAACGG	TAACCACGCA	360
GCTGGCATCC	TGACCATGGG	CCGCCCGCGCA	GGCGCAGAGC	TAGAGCCATA	TCCCTGCCCT	420
GGTCGCCGCT	GTCCGACTGC	AACCGCCACC	GCTTTAGCGC	CCCGGGGCCGG	ATCCAGAGTC	480
TGAACCGTC	TTCTATCCCT	GTCCTAGTCC	TAACTTCCC	CTCTCCTCGC	CGGTCCCTAG	540
GCAATAAAGA	CGTTTCTCTG	CTAAAAAAA	AAAAAAA	AAAAA		585

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Asn	Leu	Pro	Ser	Thr	Lys	Val	Pro	Trp	Ala	Ala	Val	Thr	Leu	Leu
1					5					10			15		
Leu	Leu	Leu	Leu	Leu	Pro	Pro	Ala	Leu	Leu	Ser	Leu	Gly	Val	Asp	Ala
					20					25			30		
Gln	Pro	Leu	Pro	Asp	Cys	Cys	Arg	Gln	Lys	Thr	Cys	Ser	Cys	Arg	Leu
					35				40			45			
Tyr	Glu	Leu	Leu	His	Gly	Ala	Gly	Asn	His	Ala	Ala	Gly	Ile	Leu	Thr
					50				55			60			

Leu Gly Lys Arg Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln
65 70 75 80
Arg Leu Leu Gln Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
85 90 95
Gly Arg Arg Ala Gly Ala Glu Leu Glu Pro Tyr Pro Cys Pro Gly Arg
100 105 110
Arg Cys Pro Thr Ala Thr Ala Leu Ala Pro Arg Gly Gly Ser
115 120 125
Arg Val
130

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
1 5 10 15
Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
20 25 30

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
1 5 10 15
Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
20 25 30
Leu

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln Ala
1 5 10 15
Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
20 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Val	Pro	Trp	Ala	Ala	Val	Thr	Leu	Leu	Leu	Leu	Leu	Leu	Pro	Pro	
1					5				10				15		
Ala	Leu	Leu	Ser	Leu	Gly	Val	Asp	Ala	Gln	Pro	Leu	Pro	Asp	Cys	Cys
						20			25				30		
Arg	Gln	Lys	Thr	Cys	Ser	Cys	Arg	Leu	Tyr	Glu	Leu	Leu	His	Gly	Ala
						35			40				45		
Gly	Asn	His	Ala	Ala	Gly	Ile	Leu	Thr	Leu	Gly	Lys	Arg	Arg	Pro	Gly
						50			55			60			
Pro	Pro	Gly	Leu	Gln	Gly	Arg	Leu	Gln	Arg	Leu	Leu	Gln	Ala	Asn	Gly
						65			70			75		80	
Asn	His	Ala	Ala	Gly	Ile	Leu	Thr	Met	Gly	Arg	Arg	Ala	Gly	Ala	Glu
						85			90			95			
Leu	Glu	Pro	His	Pro	Cys	Ser	Gly	Arg	Gly	Cys	Pro	Thr	Val	Thr	Thr
						100			105			110			
Thr	Ala	Leu	Ala	Pro	Arg	Gly	Gly	Ser	Gly	Val					
						115			120						

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
1 5 10 15
Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
20 25 30
Leu

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln
1 5 10 15
Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
20 25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CAACCNCTNC CNGACTGCTG

20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATNCCNGCNG CATGATT

17

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTGCCAGCT CCGTGCAACA GTTCGTAGAG ACGG

34

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGCAGGAAC ACGTCTTCTG GCG

23

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCCTTGGGTA TTTGGACCAC TGCACCGAAG

30

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATACCATCTC TCCGGATTGC CTCTCCCTGA

30

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCTCTGAAGG TTCCAGAATC GATAGTAN

28

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCTCTGAAGG TTCCAGAATC GATAG

25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 577 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CACAATTGAC AGCCTCAAGG TTCCTGGCTT TTTGAACCAC CACAGACATC TCCTTTCCCG 60
GCTACCCAC CCTGAGCGCC AGACACCAG AACCTTCCTT CCACAAAGGT CTCCTGGGCC 120
GCCGTGACGC TACTGCTGCT GCTGCTGCTG CTGCCGCCCG CGCTGTTGTC GTCCGGGCG 180
GCTGCACAGC CCCTGCCCGA CTGCTGTCGT CAAAAGACTT GCTCTGCCG CCTCTACGAG 240
CTGCTGCACG GCGCGGGCAA TCACGCGGCC GGCACTCCTCA CGCTGGGCAA GCGGAGGTCC 300
GGGCCCCCGG GCCTCCAGGG TCGGCTGCAG CGCCTCCTGC AGGCCAGCGG CAACCACGCC 360
GCGGGCATCC TGACCATGGG CCGCCGCGCA GGCGCAGAGC CAGCGCCGCG CCCCTGCCTC 420
GGGCGCCGCT GTTCCGCCCG GGCCGCCGCC TCCGTCGCGC CCGGAGGACA GTCCGGGATC 480
TGAGTCGTTG TTCGGGCCCT GTCCTGGCCC AGGCCTCTGC CCTCTGCCA CCCAGCGTCA 540
GCCCCAGAA AAAAGGCAAT AAAGACGAGT CTCCATT 577

cont